

# Mopping Up Metadata

Excel Functions, Formatting, Formulae

AABC-ARMA VI “Dominion of the Digital” Conference  
Victoria, BC, 28 April 2017

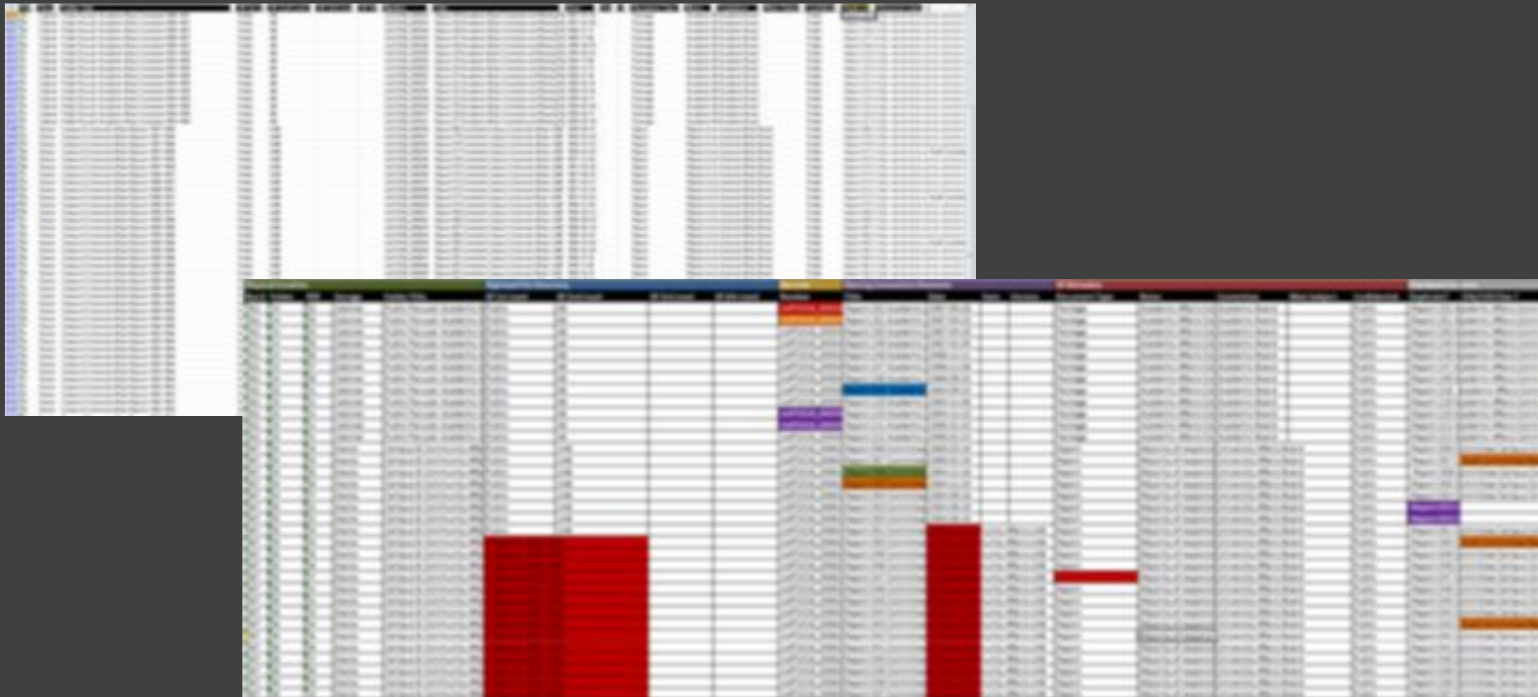
Katrina Cohen-Palacios, Adjunct Archivist  
Clara Thomas Archives and Special Collections  
York University Libraries

# map

- » **formatting**, conditional
- » **formulae**: concatenate; len; countif; today, date, year
- » **functions**: filters, sort, data validation

# conditional formatting

Difficult to spot errors in data input without conditional formatting



The image displays two side-by-side screenshots of a spreadsheet. The top screenshot shows a standard spreadsheet with no formatting. The bottom screenshot shows the same spreadsheet with conditional formatting applied, highlighting errors in red, blue, purple, and orange. The errors are categorized as follows: double spaces (blue), duplicate data (purple), and character limits (orange). The red highlights indicate data that exceeds the character limit.

Conditional formatting visualizes, highlights and categorizes error, including invisible errors such as double spaces (in blue), duplicate data (in purple), and character limits (in orange).

# conditional formatting

- » visualize data
- » highlight information
- » customize with formulae



## potential applications

- » mark required fields in red
- » highlight common errors (restricted characters, double spaces, space at end), duplicates, etc.

# conditional formatting

## » double spaces

► Use a formula to determine which cells to format

Edit the Rule Description:

**Format values where this formula is true:**

`=COUNTIF(K1,"* *")`

**Preview:** AaBbCcYyZz

Format...

## » leading spaces

► Use a formula to determine which cells to format

Edit the Rule Description:

**Format values where this formula is true:**

`=COUNTIF(K1,"* ")`

**Preview:** AaBbCcYyZz

Format...

## » trailing spaces

► Use a formula to determine which cells to format

Edit the Rule Description:

**Format values where this formula is true:**

`=COUNTIF(K1,"* ")`

**Preview:** AaBbCcYyZz

Format...

## » character limits

► Use a formula to determine which cells to format

Edit the Rule Description:

**Format values where this formula is true:**

`=LEN(V172)>=244`

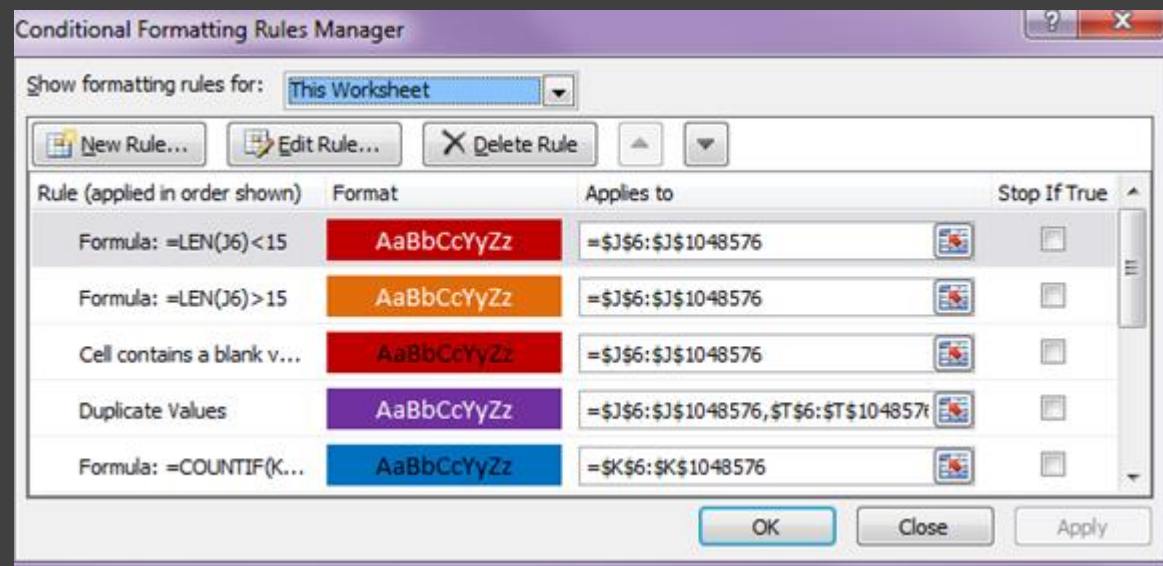
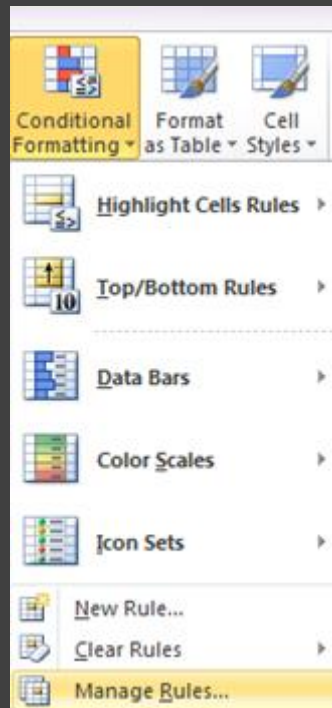
**Preview:** AaBbCcYyZz

Format...

# conditional formatting

## rules manager

» set rule precedence



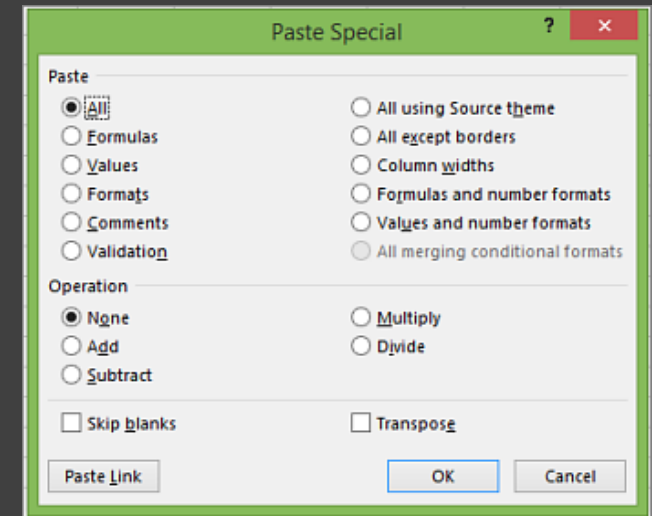
# conditional formatting

## pitfalls

» conditional formatting removed when pasting information to cells

» solutions:

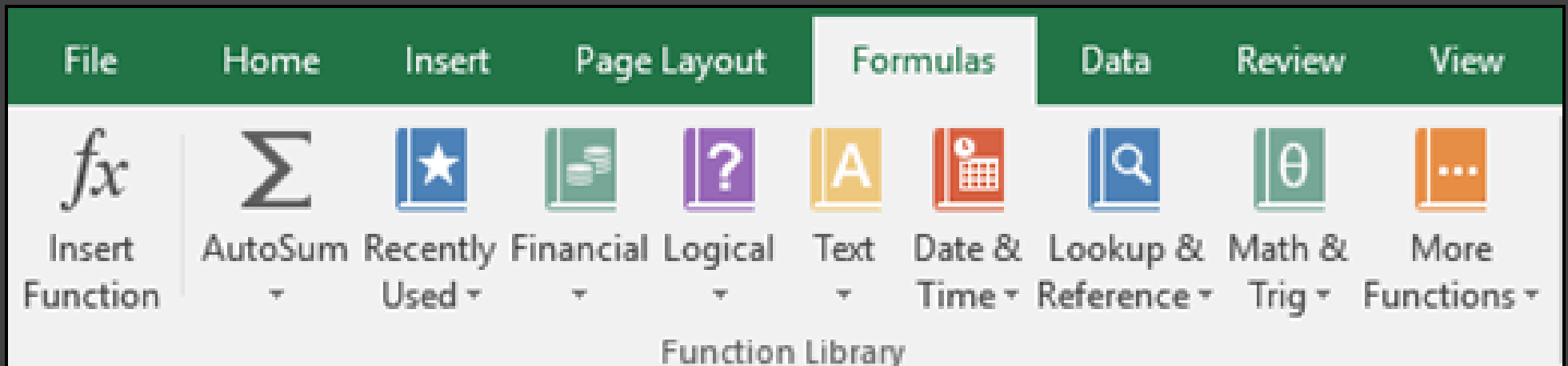
1. **special paste** values into cell & leave first row blank to special paste formatting to completed data
2. use **rules manager** to visually verify the application of formatting to a set of rows and/or columns (see previous slide)



# formulae

- » concatenate
- » len
- » countif
- » today, date, year

To refer to	Use
The cell in column A and row 10	A10
The range of cells in column A and rows 10 through 20	A10:A20
The range of cells in row 15 and columns B through E	B15:E15
All cells in row 5	5:5
All cells in rows 5 through 10	5:10
All cells in column H	H:H
All cells in columns H through J	H:J
The range of cells in columns A through E and rows 10 through 20	A10:E20





# formulae » concatenate

» strings two or more text into one string

**=CONCATENATE (A1, "text", B1, ...)**

» use quotation marks to insert text

» concatenate replaced by concat in Excel 2016

Notes	Script	Signatures	Archival Accession #	
=CONCATENATE(N6," that contains ",O6," and is located in under accession number: ",P6)	Manuscript	signatures	A19	1/167
M CONCATENATE(text1, [text2], [text3], [text4], [text5], [text6], ...)				
signatures; and is located in under accession number: A19 1/167	Manuscript	signatures	A19	1/167
Manuscript that contains signatures, and is located in under accession number: A19 1/167	Manuscript	signatures	A19	1/167

# formulae » concatenate » uses

## potential applications

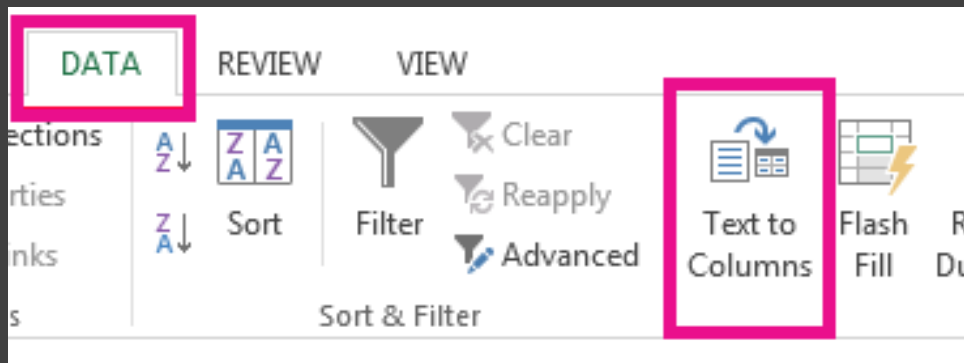
- » determine characters used with LENS formula
- » automate fields and standardized notes

## pitfalls

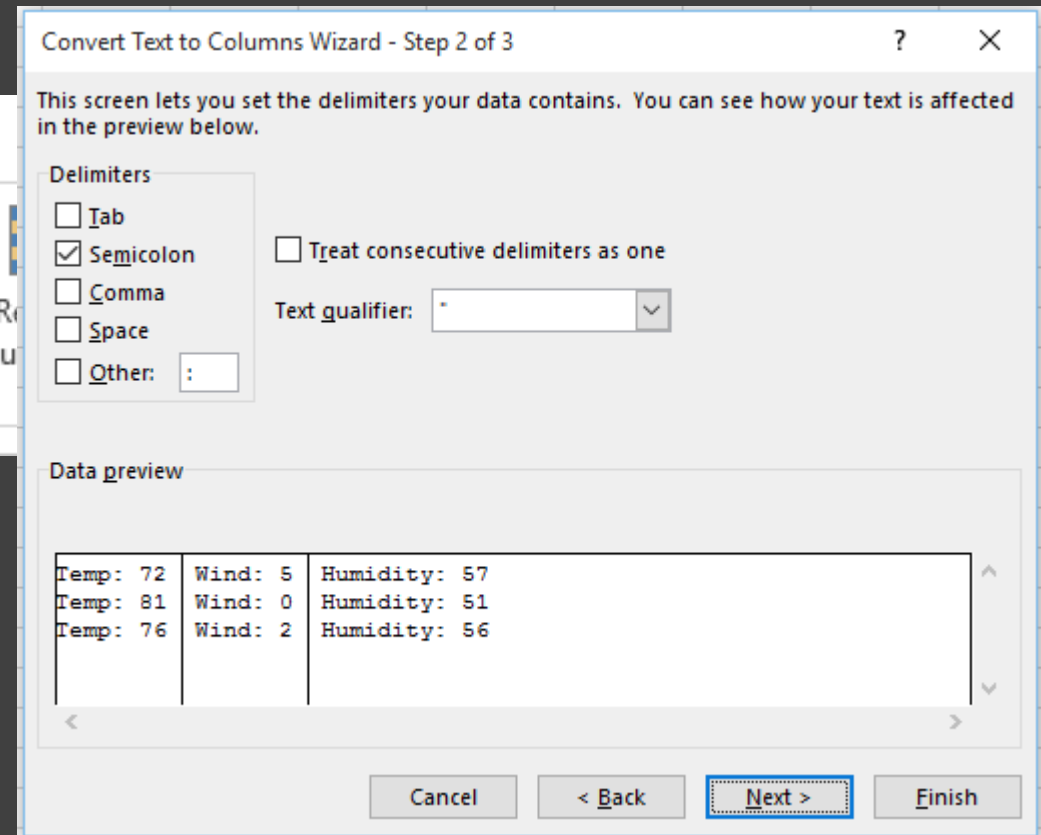
- » if referenced cells are deleted, the information is also deleted from formulated cell
- » prior to deleting referenced cell, copy and special paste values in formulated cell

# formulae » concatenate » spill cells

» opposite: “text to columns” option on “data” tab



Images from: <https://support.office.com/en-us/article/Split-a-cell-f1804d0c-e180-4ed0-a2ae-973a0b7c6a23>



# formulae » len

- » returns the number of characters in a text string

**=LEN(A1+B1)**

## pitfall

- » number and date formatting is not included in count  
(ie: 100 formatted as \$100.00 returns 3)

## potential applications

- » calculate characters based on restrictions

# formulae » len

File Name Controls		
Duplicate?	Path	256/1
-Report 223 A	Public-AB---Report 22	86
-Report 224 A	http://servername.uta	76
-Report 225 A	http://servername.uta	76
-Report 226 A	http://servername.uta	76
-Report 227 A	http://servername.uta	=LEN(V10)
-Report 080 C	http://servername.uta	LEN(text)
-Report 079 C	http://servername.uta	//

File Name Controls		
Duplicate?	Path	256/1
-Report 223 A	Public-AB---Report 22	86
-Report 224 A	http://servername.uta	76
-Report 225 A	http://servername.uta	76
-Report 226 A	=CONCATENATE("http:",G9,"-",H9,"-",I9,"-",J9,"-",U9)	
-Report 227 A	CONCATENATE(text1, [text2], [text3], [text4], [text5], [text6], [text7], [text8], [text9], [text10], [text11], ...)	
-Report 080 C	http://servername.uta	//

LEN formula combined with CONCATENATE to determine character limits for a file path in Sharepoint

# formulae » **countif**

- » calculates cells that meet a criterion

**=COUNTIF(A2:A99, "text")**

**=COUNTIF(A2:A99, A1)**

- » can use wildcard characters (?,\*) in criteria
- » remove leading spaces, trailing spaces, etc
- » **COUNTA** counts cells that are not empty
- » **COUNTBLANK** counts empty cells

## formulae » today, year, day

- » =TODAY() returns the current date
- » =TODAY()+5 returns date plus five days
- » =YEAR(TODAY()) returns year of current date
- » =MONTH(TODAY()) returns month of date
- » =DAY(TODAY()) returns day of current month

**potential applications:** calculate anniversaries

**=YEAR((TODAY())-YEAR(C1))**

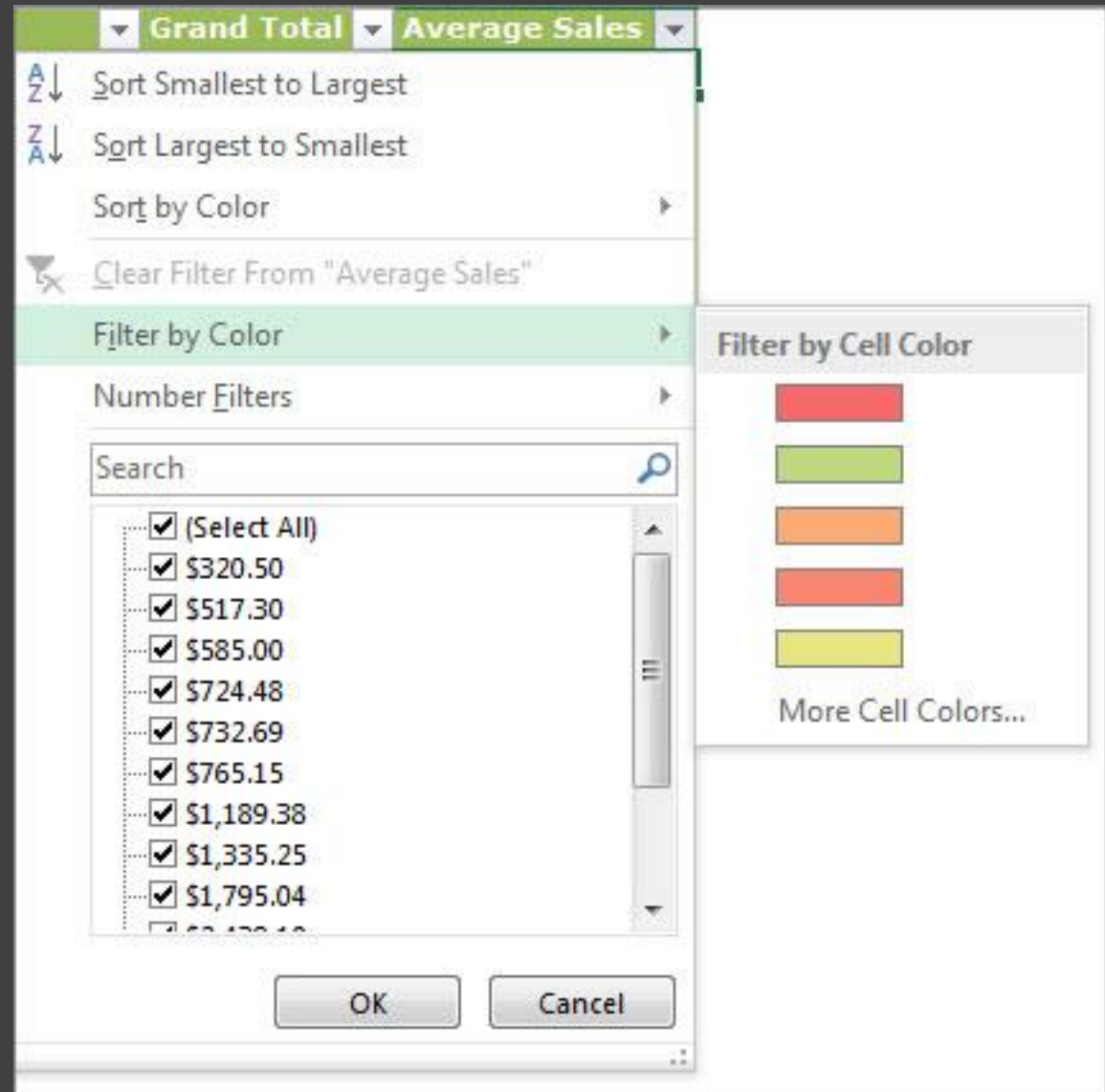
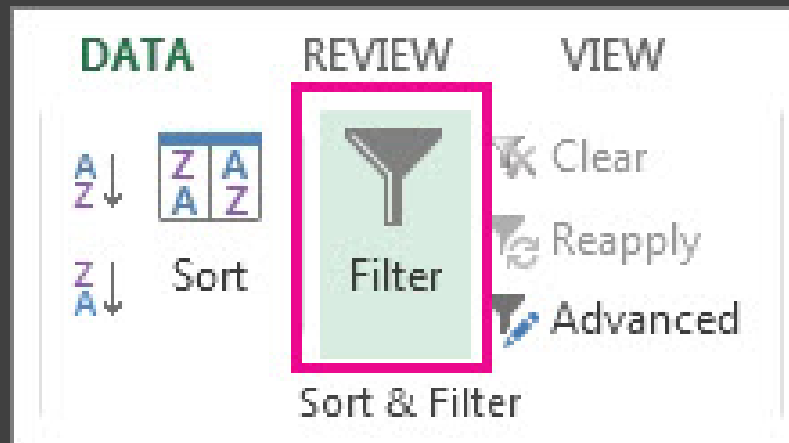
**pitfall:** cannot calculate dates prior to 1900

# functions

- » filters
- » sort
- » data validation



# filter



# filter

hide unwanted data (but does not delete it)

- » auto-filters: by **value**, by **format**, or by **criteria**
- » each filter type is mutually exclusive
- » customize filters

**common error:** ensure all columns and rows are included in the filtered data

**tip:** use freeze pane to view header columns when scrolling; boolean searching

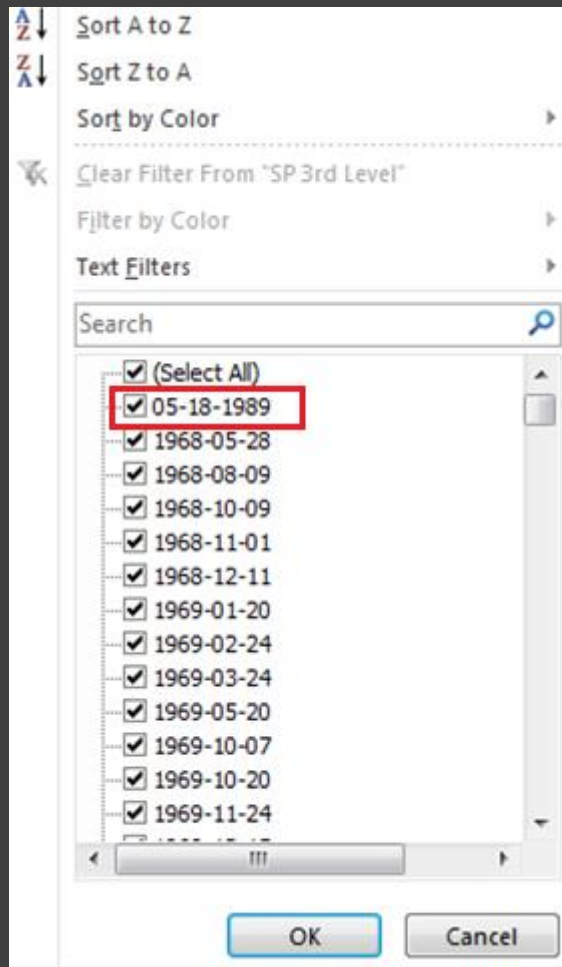
# filter » uses

## potential applications

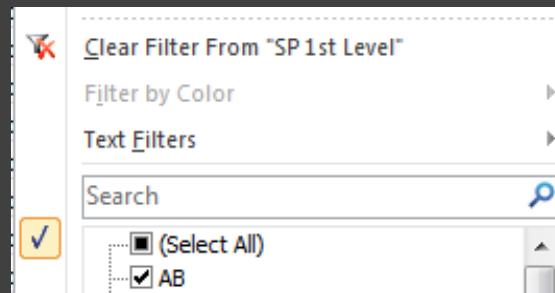
- » verify consistency/format of data
- » highlights misspellings
- » used to inventory vendor outputs  
ex: # of rows = # of PDFs returned
- » view upcoming anniversary dates

**pitfall:** filtered out data is deleted  
when file is saved as a CSV

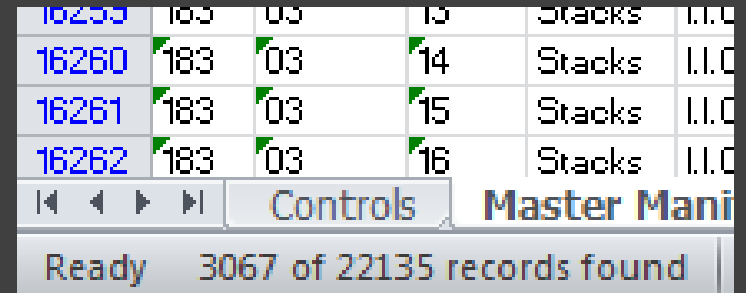
# filter



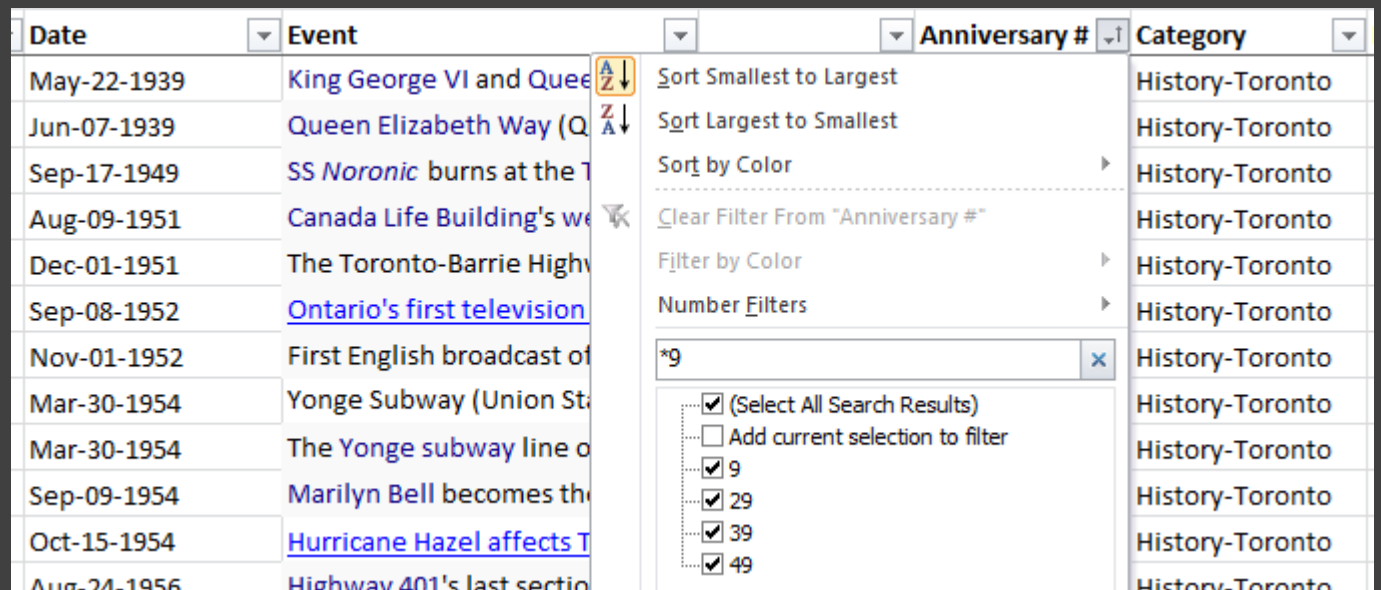
Quickly spot inconsistent data formatting



Hide groups of data with filters

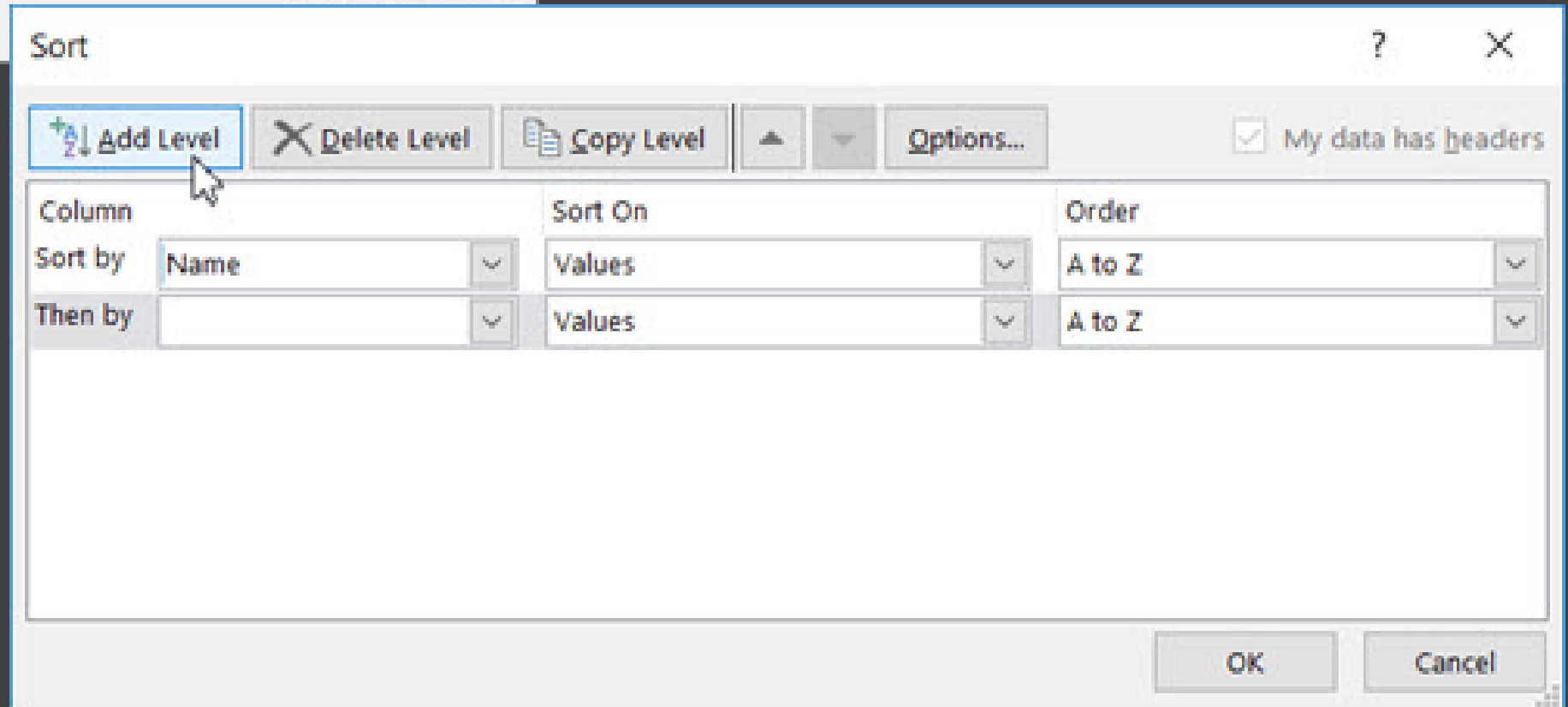
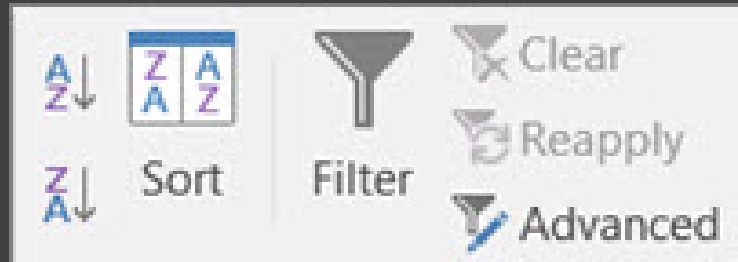


Inventory and counts data through filters



Discover and calculate data such as upcoming anniversaries for social media posts

# sort



# sort

## rearrange data

- » by **text**: A to Z or reverse
- » by **numbers**: smallest to largest or reverse
- » by **date/time**: oldest to newest or reverse

**common errors**: remove leading spaces, verify if all cells are formatted text or numbers, sorting does not move hidden rows

## sort » uses

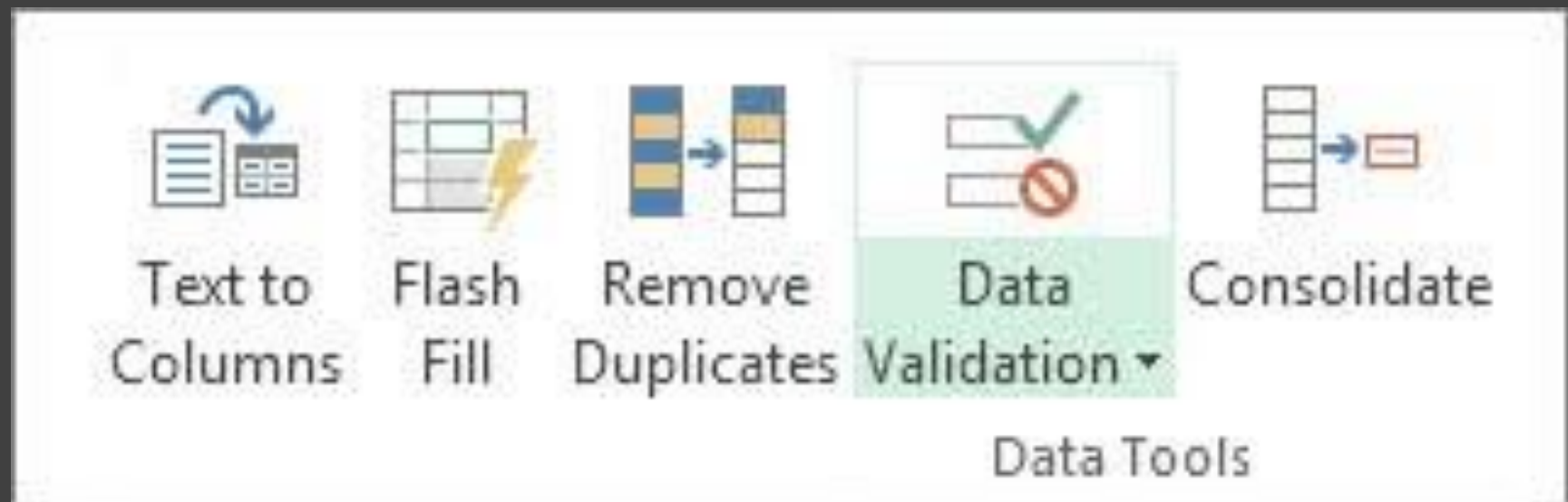
### potential applications

- » rearrange data to reflect original order by sorting by box, file, and item numbers
- » review entire series and maintain original order
- » spot inconsistencies in data input

# data validation

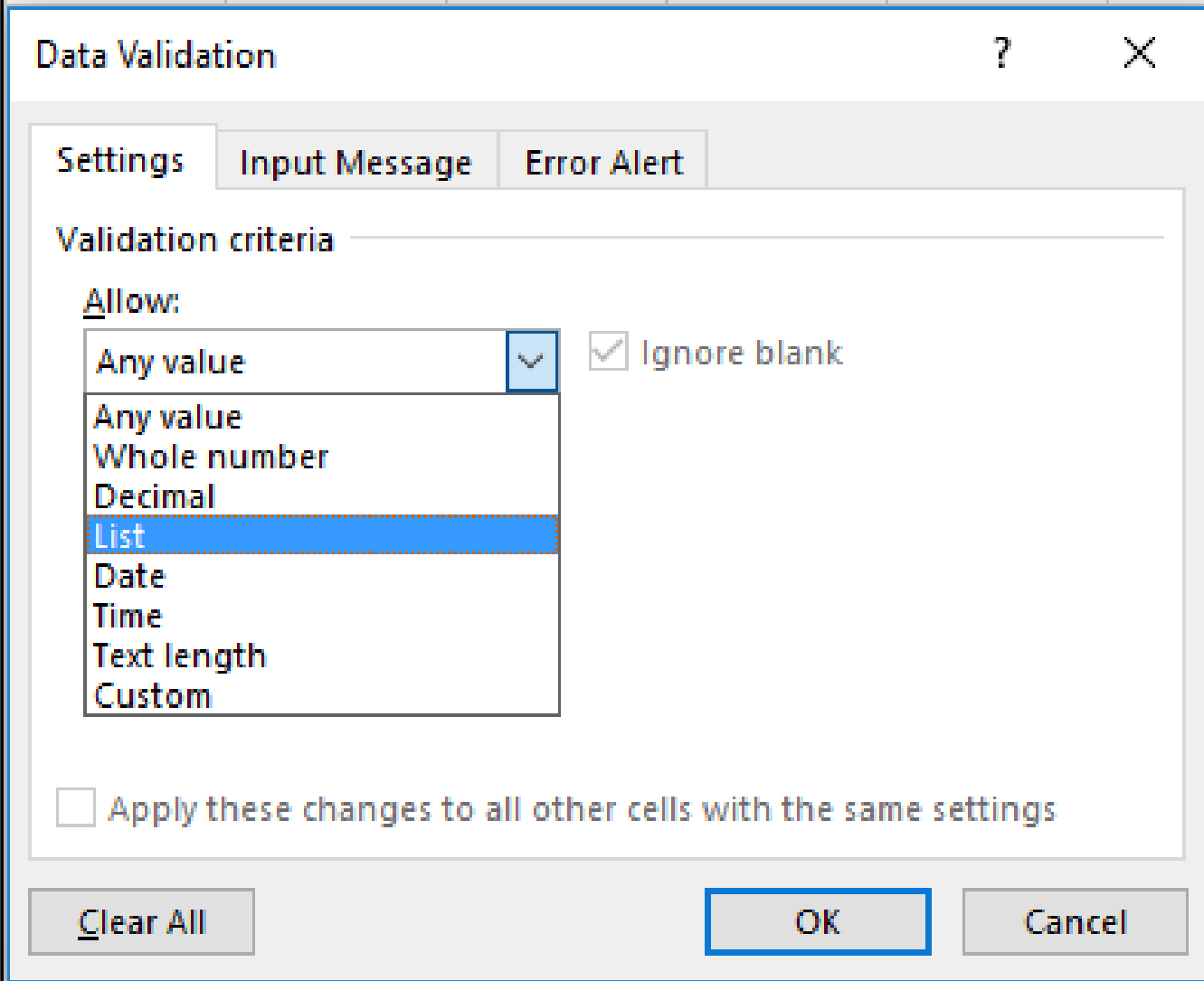
restrict type of data/values entered in a cell

- » set up a control list in a worksheet and select the entire column to ensure additions/deletions are updated automatically the validation list





# data validation » settings



The screenshot shows the 'Data Validation' dialog box with the 'Settings' tab selected. The 'Validation criteria' section is visible, showing a dropdown menu for 'Allow:' with 'List' selected. The 'Ignore blank' checkbox is checked. At the bottom, there is a checkbox for 'Apply these changes to all other cells with the same settings' which is unchecked. The 'Clear All', 'OK', and 'Cancel' buttons are at the bottom of the dialog.

Data Validation

Settings Input Message Error Alert

Validation criteria

Allow:

Any value  
Any value  
Whole number  
Decimal  
List  
Date  
Time  
Text length  
Custom

☒ Ignore blank

☐ Apply these changes to all other cells with the same settings

Clear All OK Cancel

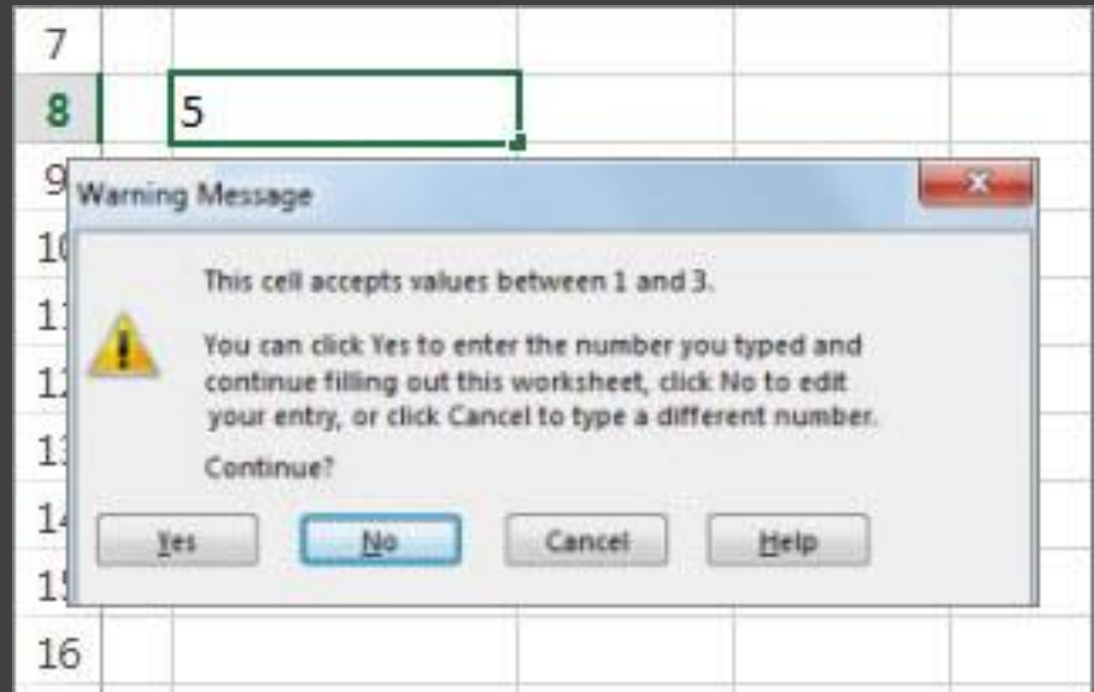
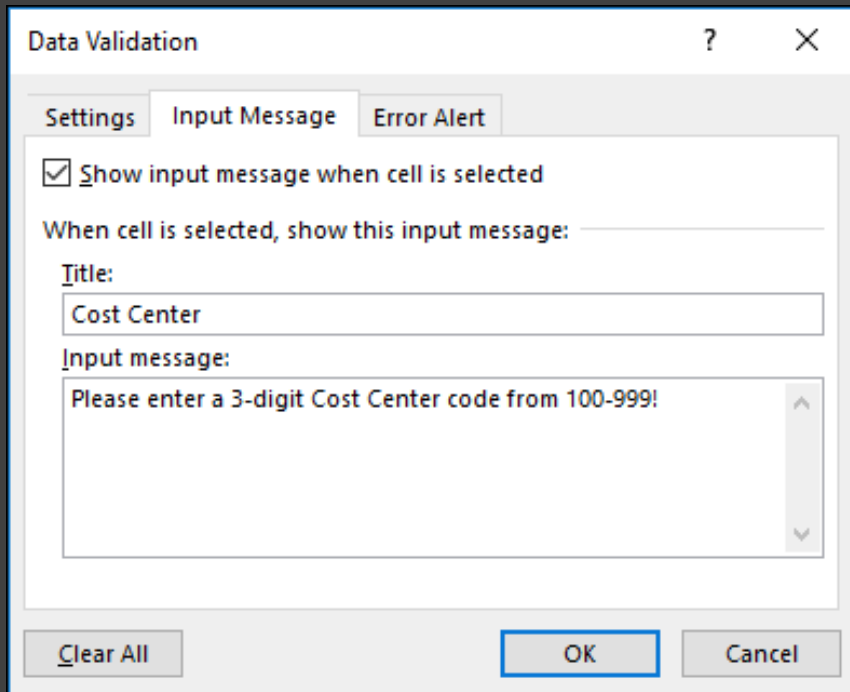
## data validation » settings

**error alert** appears if user enters invalid data

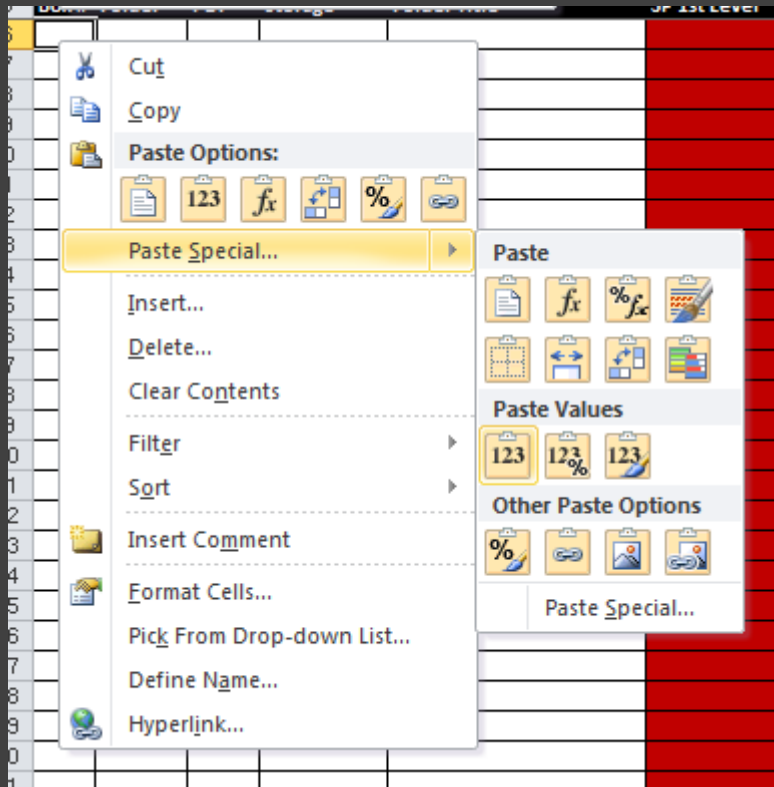
- » **stop**: user must retry or cancel data input
- » **warning**: warning appears but does not prevent
- » **information**: informs users data is invalid

# data validation » settings

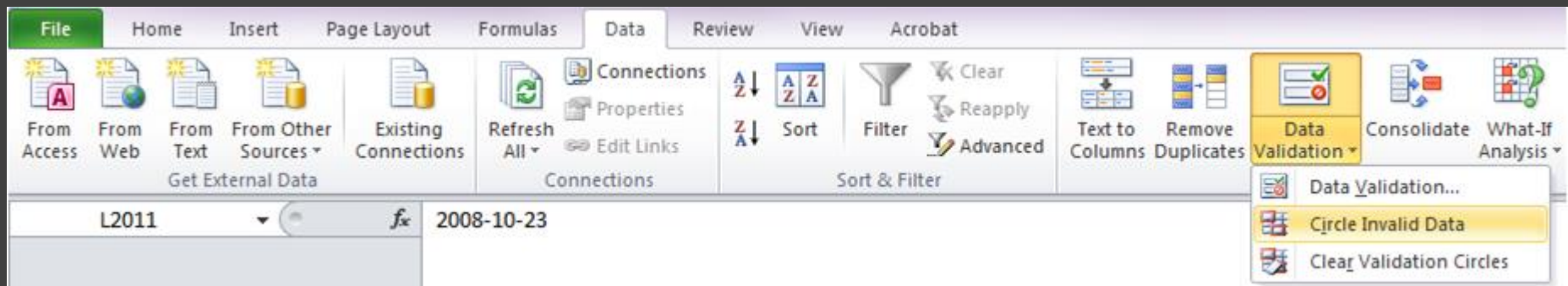
input message: define instructions



# data validation » pitfalls



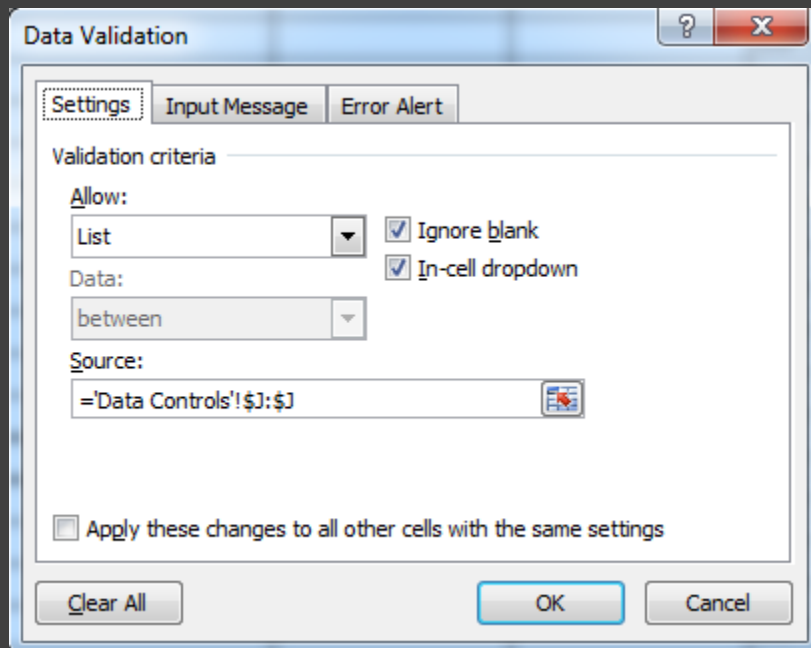
- » pasting data removes data validation
- » paste values only via “special paste” option
- » error alerts are not applied retroactively
- » circle invalid data to review



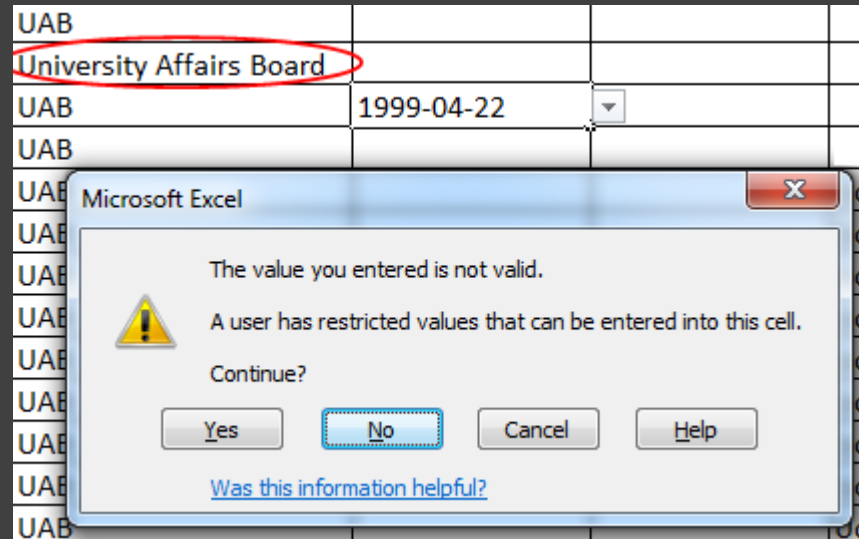
# data validation

SP 2nd Level	SP 3rd Level	SP 4th
AB		
AB		
AB	1940-1949	
AB	1950-1959	
AB	1960-1969	
AB	1970-1979	
UAB	1980-1989	
UAB	1990-1999	
UAB	2000-2009	
UAB	2010-2015	

Data validation creates drop down list for input



Data validation uses a control list from the spreadsheet file

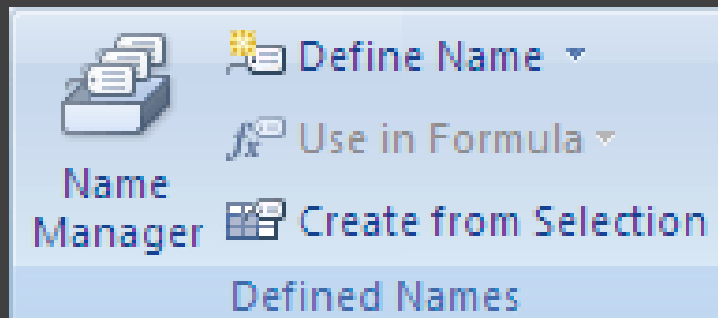


Data validation circles invalid data and creates input warnings

# data validation » pitfalls

## multiple users

- » create control list as a separate workbook
- » design name for each list from formulas tab
- » in data validation workbook, add the list via the name manager and enter address in “refers to” section of data validation



# questions?

## conditional formatting

» visualize data

**filter** » hide data

**sort** » rearrange data

## data validation

» control data entered in cell

## formulae

=CONCATENATE (A1, "text")

=LEN(A1+B1)

=COUNTIF(A2:A99,"TEXT")

=YEAR((TODAY()))-YEAR(C1))

## contact info

kcohenp@yorku.ca

twitter: @katrina\_cp